Collaborative Large Language Model for Recommender Systems

Abstract

Recently, there has been growing interest in developing the next-generation recommender systems (RSs) based on pretrained large language models (LLMs). However, the semantic gap between natural language and recommendation tasks is still not well addressed, leading to multiple issues such as spuriously correlated user/item descriptors, ineffective language modeling on user/item data, inefficient recommendations via auto-regression, etc. In this paper, we propose CLLM4Rec, the first generative RS that tightly integrates the LLM paradigm and ID paradigm of RSs, aiming to address the above challenges simultaneously. We first extend the vocabulary of pretrained LLMs with user/item ID tokens to faithfully model user/item collaborative and content semantics. Accordingly, a novel soft+hard prompting strategy is proposed to effectively learn user/item collaborative/content token embeddings via language modeling on RS-specific corpora, where each document is split into a prompt consisting of heterogeneous soft (user/item) tokens and hard (vocab) tokens and a main text consisting of homogeneous item tokens or vocab tokens to facilitate stable and effective language modeling. In addition, a novel mutual regularization strategy is introduced to encourage CLLM4Rec to capture recommendation-related information from noisy user/item content. Finally, we propose a novel recommendation-oriented finetuning strategy for CLLM4Rec, where an item prediction head with multinomial likelihood is added to the pretrained CLLM4Rec backbone to predict hold-out items based on soft+hard prompts established from masked user-item interaction history, where recommendations of multiple items can be generated efficiently without hallucination.